# IT & C

ISSN 2821 - 8469, ISSN – L 2821 - 8469, Volumul 1, Numărul 1, Septembrie 2022

# Cercetare și aplicații în rețelele sociale

Bogdan Batrinca, Philip C. Treleaven

**Pentru a cita acest articol**: Batrinca, Bogdan și Treleaven, Philip C. (2022), Cercetare și aplicații în rețelele sociale, *IT & C*, **1**:1, 63-68, https://www.internetmobile.ro/cercetare-si-aplicatii-in-retelele-sociale/. Traducere și adaptare independente Nicolae Sfetcu

Publicat online: 21.08.2022

## ABONARE

# Cercetare și aplicații în rețelele sociale

Bogdan Batrinca, Philip C. Treleaven

**Rezumat**

Rețelele sociale sunt definite ca aplicații de internet bazate pe web și pe mobil care permit crearea, accesul și schimbul de conținut generat de utilizatori care este accesibil omniprezent. Pe lângă rețelele sociale (de exemplu, Twitter și Facebook), pentru comoditate, vom folosi termenul „social media” pentru a include și fluxuri RSS, bloguri, wiki-uri și știri, toate generând de obicei text nestructurat și accesibile prin intermediul web. Rețelele sociale sunt importante în special pentru cercetarea în știința socială computațională care investighează întrebări folosind tehnici cantitative (de exemplu, statistici computaționale, învățarea automată și complexitate) și așa-numitele megadate (big data) pentru extragerea datelor și modelarea prin simulare.

Sursa: Bogdan Batrinca, Philip C. Treleaven, "Social media analytics: a survey of techniques, tools and platforms", *AI & Soc* (2015) 30:89-116 DOI 10.1007/s00146-014-0549-4, Creative Commons Attribution License. Traducere și adaptare independente: Nicolae Sfetcu

**Cuvinte cheie**: cercetare, aplicații, rețele sociale

IT & C, Volumul 1, Numărul 1, Septembrie 2022, pp. 63-68
ISSN 2821 - 8469, ISSN – L 2821 - 8469
URL: https://www.internetmobile.ro/cercetare-si-aplicatii-in-retelele-sociale/




Rețelele sociale sunt definite ca aplicații de internet bazate pe web și pe mobil care permit crearea, accesul și schimbul de conținut generat de utilizatori care este accesibil omniprezent (Kaplan și Haenlein 2010). Pe lângă rețelele sociale (de exemplu, Twitter și Facebook), pentru comoditate, vom folosi termenul „social media” pentru a include și fluxuri RSS, bloguri, wiki-uri și știri, toate generând de obicei text nestructurat și accesibile prin intermediul web. Rețelele sociale sunt importante în special pentru cercetarea în știința socială computațională care investighează întrebări (Lazer și colab. 2009) folosind tehnici cantitative (de exemplu, statistici computaționale, învățarea automată și complexitate) și așa-numitele megadate (big data) pentru extragerea datelor și modelarea prin simulare (Cioffi- Revilla 2010).

Acest lucru a condus la numeroase servicii de date, instrumente și platforme de analiză. Cu toate acestea, această disponibilitate ușoară a datelor din rețelele sociale pentru cercetarea academică se poate schimba semnificativ din cauza presiunilor comerciale. În plus, instrumentele disponibile cercetătorilor sunt departe de a fi ideale. Acestea fie oferă acces superficial la datele

brute, fie (pentru acces nesuperficial) solicită cercetătorilor să programeze analitice într-un limbaj precum Java.

## Terminologie

Definițiile unora dintre tehnicile cheie legate de analiza datelor textuale nestructurate:

- **Procesarea limbajului natural** — (natural language processing, NLP) este un domeniu de informatică, inteligență artificială și lingvistică preocupat de interacțiunile dintre computere și limbajele umane (naturale). Mai exact, este procesul unui computer care extrage informații semnificative din intrarea în limbaj natural și/sau produce ieșiri în limbaj natural.
- **Analitica știrilor** — măsurarea diferitelor atribute calitative și cantitative ale știrilor (date nestructurate). Unele dintre aceste atribute sunt: sentimentul, relevanța și noutatea.
- **Mineritul opiniilor** — mineritul opiniei (mineritul sentimentelor, extracția opiniei/sentimentului) este domeniul de cercetare care încearcă să creeze sisteme automate pentru a determina opinia umană din text scris în limbaj natural.
- **Scraping**—colectarea de date online de pe rețelele sociale și alte site-uri web sub formă de text nestructurat și cunoscută și sub denumirea de site scraping, recoltare web și extragere de date web.
- **Analiza sentimentelor** — analiza sentimentelor se referă la aplicarea procesării limbajului natural, a lingvisticii computaționale și a analizei de text pentru a identifica și extrage informații subiective din materialele sursă.
- **Analitica textelor** — implică extragerea informațiilor (information retrieval, IR), analiza lexicală pentru a studia distribuțiile de frecvență a cuvintelor, recunoașterea modelelor, etichetarea/adnotarea, extragerea informațiilor, tehnicile de minerit a datelor, inclusiv analiza de legături și asocieri, vizualizare și analitica predictivă.

## Provocări în cercetare

Scraping (scanarea) și analitica rețelelor sociale oferă o sursă bogată de provocări de cercetare academică pentru oamenii din științe sociale, informaticieni și organismele de finanțare. Provocările includ:

- **Scraping** — deși datele din rețelele sociale sunt accesibile prin intermediul API-urilor, din cauza valorii comerciale a datelor, majoritatea surselor majore, cum ar fi Facebook și Google, fac din ce în ce mai dificil pentru cadrele universitare să obțină acces complet la datele lor „brute”; foarte puține surse de date sociale oferă date la prețuri accesibile mediului academic și cercetătorilor. Serviciile de știri cum ar fi Thomson Reuters și Bloomberg percep de obicei o sumă pentru accesul la datele lor. În schimb, Twitter a anunțat recent programul Twitter Data Grants, în care cercetătorii pot aplica pentru a obține acces la tweet-urile publice și datele istorice ale lui Twitter pentru a obține informații din setul său masiv de date (Twitter are mai mult de 500 de milioane de tweet-uri pe zi).

- **Curățirea datelor** — curățarea datelor textuale nestructurate (de exemplu, normalizarea textului), în special a datelor transmise în flux de înaltă frecvență în timp real, prezintă încă numeroase probleme și provocări de cercetare.
- **Surse de date holistice** — cercetătorii reunesc și combină din ce în ce mai mult surse de date noi: date de pe rețelele sociale, date de piață și clienți în timp real și date geospațiale pentru analiză.
- **Protecția datelor** — odată ce ați creat o resursă de „megadate" ("big data"), datele trebuie securizate, problemele de proprietate și de IP rezolvate (adică, stocarea datelor extrase este împotriva majorității termenilor de serviciu ai editorilor), iar utilizatorii să aibă la dispoziție diferite niveluri de acces; în caz contrar, utilizatorii pot încerca să „sugă" toate datele valoroase din baza de date.
- **Analitica datelor** — analiza sofisticată a datelor din rețelele sociale pentru mineritul opiniilor (de exemplu, analiza sentimentelor) ridică încă o multitudine de provocări din cauza limbilor străine, cuvintelor străine, argoului, erorilor de ortografie și evoluției naturale a limbilor.
- **Tablouri de bord pentru analitica** — multe platforme de rețele sociale impun utilizatorilor să scrie API-uri pentru a accesa fluxuri sau modele de analitica programelor într-un limbaj de programare, cum ar fi Java. Deși sunt rezonabile pentru informaticieni, aceste abilități depășesc de obicei majoritatea cercetătorilor (științe sociale). Sunt necesare interfețe care nu sunt de programare pentru a oferi ceea ce s-ar putea numi acces „profund" la datele „brute", de exemplu, configurarea API-urilor, îmbinarea fluxurilor de rețele sociale, combinarea surselor holistice și dezvoltarea modelelor analitice.
- **Vizualizarea datelor** — reprezentare vizuală a datelor prin care informațiile care au fost extrase într-o formă schematică cu scopul de a comunica informațiile în mod clar și eficient prin mijloace grafice. Având în vedere amploarea datelor implicate, vizualizarea devine din ce în ce mai importantă.

## Cercetare și aplicații în rețelele sociale

Datele din rețelele sociale sunt în mod clar cea mai mare, mai bogată și mai dinamică bază de înregistrări ale comportamentului uman, aducând noi oportunități de a înțelege indivizii, grupurile și societatea. Oamenii de știință inovatori și profesioniștii din industrie găsesc din ce în ce mai multe modalități noi de a colecta, combina și analiza automat această bogăție de date. Desigur, a face dreptate acestor aplicații de rețele sociale de pionierat în câteva paragrafe este o provocare. Trei domenii ilustrative sunt: afaceri, bioștiințe și științe sociale.

Primii care au adoptat analiza rețelelor sociale de afaceri au fost de obicei companii din retail și finanțe. Companiile de retail folosesc rețelele sociale pentru a-și valorifica gradul de cunoaștere a mărcii, îmbunătățirea produselor/serviciului pentru clienți, strategiile de publicitate/ marketing, analiza structurii rețelei, propagarea știrilor și chiar detectarea fraudei. În finanțe, rețelele de socializare sunt folosite pentru a măsura sentimentul pieței, iar datele de știri sunt folosite pentru tranzacționare. Ca o ilustrare, Bollen et al. (2011) au măsurat sentimentul unui

eșantion aleatoriu de date Twitter, constatând că prețurile Dow Jones Industrial Average (DJIA) sunt corelate cu sentimentul Twitter cu 2-3 zile mai devreme, cu o acuratețe de 87,6%. Wolfram (2010) a folosit datele Twitter pentru a antrena un model de regresie vectorială de suport (SVR) pentru a prezice prețurile acțiunilor individuale NASDAQ, găsind un „avantaj semnificativ" pentru estimarea prețurilor 15 minute în viitor.

În bioștiințe, rețelele sociale sunt folosite pentru a colecta date despre cohorte mari pentru inițiative de schimbare a comportamentului și monitorizarea impactului, cum ar fi combaterea fumatului și a obezității sau monitorizarea bolilor. Un exemplu este biologii de la Penn State University (Salathe et al. 2012) care au dezvoltat sisteme și tehnici inovatoare pentru a urmări răspândirea bolilor infecțioase, cu ajutorul site-urilor web de știri, blogurilor și rețelelor sociale.

Aplicațiile computaționale ale științelor sociale includ: monitorizarea răspunsurilor publice la anunțuri, discursuri și evenimente, în special comentarii și inițiative politice; perspective în comportamentul comunității; sondaje pe rețelele sociale ale grupurilor (greu de contactat); detectarea timpurie a evenimentelor emergente, ca în Twitter. De exemplu, Lerman et al. (2008) folosesc lingvistica computațională pentru a prezice automat impactul știrilor asupra percepției publice a candidaților politici. Yessenov și Misailovic (2009) folosesc comentariile de recenzii ale filmelor pentru a studia efectul diverselor abordări în extragerea caracteristicilor textului asupra acurateței a patru metode de învățare automată—Naive Bayes, Decision Trees, Maximum Entropie și K-Means clustering. În cele din urmă, Karabulut (2013) a constatat că Fericirea națională brută (GNH) a Facebook prezintă maxime și minime în concordanță cu evenimentele publice majore din SUA.

## Bibliografie

Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2(3):1-8

Cioffi-Revilla C (2010) Computational social science. Wiley Inter-discip Rev Comput Statistics 2(3):259-271

Kaplan AM, Haenlein M (2010) Users of the world, unite! the challenges and opportunities of social media. Bus Horiz 53(1):59-68

Karabulut Y (2013) Can Facebook predict stock market activity? SSRN eLibrary, pp 1-58. http://ssrn.com/abstract=2017099 or http://dx.doi.org/10.2139/ssrn.2017099. Accessed 2 Feb 2014

Lazer D et al (2009) Computational social science. Science 323:721-723

Lerman K, Gilder A, Dredze M, Pereira F (2008) Reading the markets: forecasting public opinion of political candidates by news analysis. In: *Proceedings of the 22nd international conference on computational linguistics* 1:473-480

Salathe M et al. (2012) Digital epidemiology. PLoS Comput Biol 8(7):1-5

Wolfram SMA (2010) Modelling the stock market using Twitter. Dissertation Master of Science thesis, School of Informatics,

Yessenov K, Misailovic S (2009) Sentiment analysis of movie review comments, pp 1-17. http://people.csail.mit.edu/kuat/courses/6. 863/report.pdf. Accessed 16 Aug 2013

Sursa: Bogdan Batrinca, Philip C. Treleaven, "Social media analytics: a survey of techniques, tools and platforms", *AI & Soc* (2015) 30:89-116 DOI 10.1007/s00146-014-0549-4, Creative Commons Attribution License. Traducere și adaptare independente: Nicolae Sfetcu